

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-049948

(43)Date of publication of application : 18.02.2000

(51)Int.Cl.

H04M 3/56

(21)Application number : 10-215824

(71)Applicant : SONY CORP

(22)Date of filing : 30.07.1998

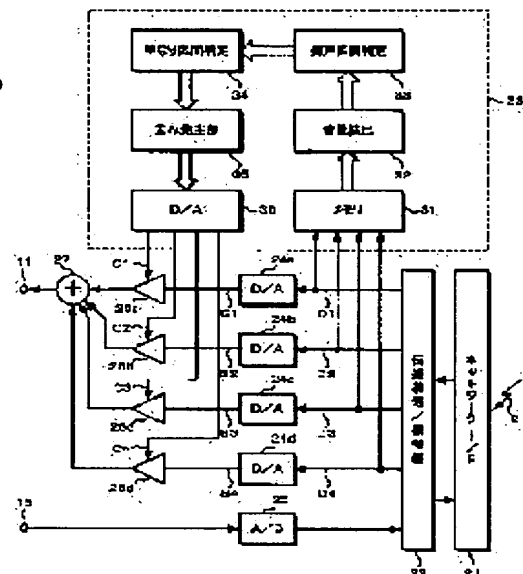
(72)Inventor : KONDO TETSUJIRO
ICHIKI HIROSHI

(54) SPEECH COMMUNICATION DEVICE AND SPEECH COMMUNICATION SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To enhance the operability of a communication system such as a telephone conference system and a speech device by facilitating to recognize the voice of an opposite party who is a center of a conversation.

SOLUTION: A plurality of communication devices are interconnected by a 2-way communication line. Each communication device is provided with a control section 23, a plurality of gain control amplifiers 26 and an adder 27. The control section 23 detects a speech time and detects an overlap time of conversation. Let an overlap time of speech be T1 and let a speech time of a person estimated to be a speech opposite party be T2, then the weight W of sound volume suppression is calculated by $W=T1/T2$. The control amplifiers are controlled by the weight W to suppress undesired voice signals and to form a pleasant communication environment.



LEGAL STATUS

[Date of request for examination]

10.03.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

BEST AVAILABLE COPY

(19)日本特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号
特開2000-49948
(P2000-49948A)

(43)公開日 平成12年2月18日(2000.2.18)

(51)IntCl.⁷
H 0 4 M 3/56

識別記号

F I
H 0 4 M 3/56

テーマコード(参考)
5 K 0 1 5

審査請求 未請求 請求項の数 5 O L (全 6 頁)

(21)出願番号 特願平10-215824

(22)出願日 平成10年7月30日(1998.7.30)

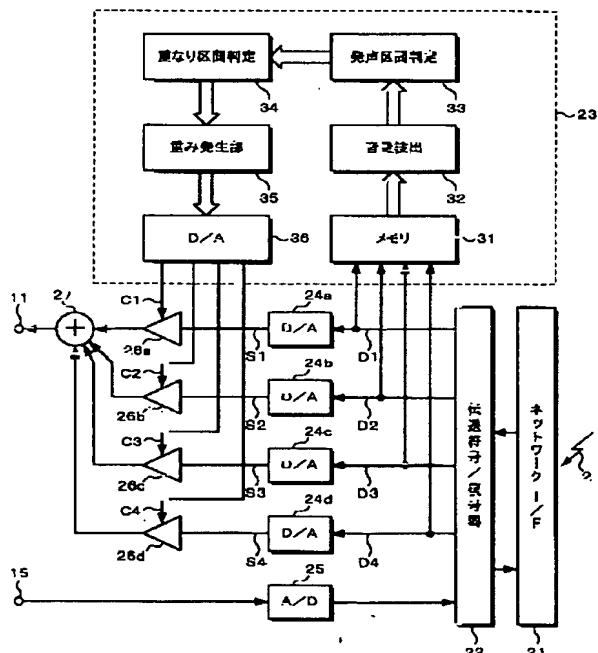
(71)出願人 000002185
ソニー株式会社
東京都品川区北品川6丁目7番35号
(72)発明者 近藤 哲二郎
東京都品川区北品川6丁目7番35号 ソニー株式会社内
(72)発明者 一木 洋
東京都品川区北品川6丁目7番35号 ソニー株式会社内
(74)代理人 100082762
弁理士 杉浦 正知
Fターム(参考) 5K015 JA05 JA10 JA11

(54)【発明の名称】 音声通話装置および音声通話システム

(57)【要約】

【課題】 電話会議システム等の通信システムおよび通話装置において、会話の中心となる相手の話の聞き分けを容易にし、使い勝手を良くする。

【解決手段】 双方向通信路を介して複数の通話装置が接続される。各通話装置に、制御部23と、話者に応じて複数設けられたゲインコントロールアンプ26と、加算器27とを設ける。制御部23において、発声時間を検出すると共に、会話の重なり時間を検出する。音声の重なり時間をT1とし、話相手と想定される人の発声時間をT2として、音量抑制の重みWを($W=T1/T2$)により算出する。重みWでゲインコントロールアンプを制御することにより、不要な音声を抑制して快適な通話環境を形成する。



【特許請求の範囲】

【請求項1】 通信を介して複数の人が音声通話するようになされた音声通話システムに用いられる音声通話装置において、

通話中に各人の音声の音量を検出する音量検出手段と、上記音量検出手段により検出される音量を所定のしきい値で2値化することにより音声の有無を判別する判別手段と、

上記判別手段の判別結果により、通話中の各人の音声の重なりを判定し、上記音声の重なりに応じて通話中の各人の音声のレベルを調整する音量調整手段とを備えたことを特徴とする音声通話装置。

【請求項2】 通信を介して複数の人が音声通話するようになされた音声通話システムにおいて、

双方向通信路を介して複数の音声通話装置が結合され、上記音声通話装置は、他の音声通話装置との通信のための通信手段と、通話中に各人の音声の音量を検出する音量検出手段と、上記音量検出手段により検出される音量を所定のしきい値で2値化することにより音声の有無を判別する判別手段と、

上記判別手段の判別結果により、通話中の各人の音声の重なりを判定し、上記音声の重なりに応じて通話中の各人の音声のレベルを調整する音量調整手段とを備えたことを特徴とする音声通話システム。

【請求項3】 請求項1または2において、上記音量調整手段は、過去の所定区間の音声信号に基づいて音量調整を行うことを特徴とする装置。

【請求項4】 請求項1または2において、上記音量調整手段の音声の重なりに応じた処理は、所定区間の音声の重なり時間と話相手と想定される人の発声時間とによる演算結果に基づいてなされることを特徴とする装置。

【請求項5】 請求項1または2において、上記音量調整手段の音声の重なりに応じた処理に用いられる演算結果は、音声の重なり時間を T_1 とし、話相手と想定される人の発声時間を T_2 として、下式により算出される W であることを特徴とする装置。

$$W = T_1 / T_2$$

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、例えば、3人以上の話者の間において音声通話する際に用いられ、電話会議や遠隔テレビ会議等に用いて好適な音声通話装置および音声通話システムに関する。

【0002】

【従来の技術】コミュニケーションを目的とし、遠隔地に音声や映像を伝送することが可能な各種装置およびシステムが提案されているが、実際に広く普及しているのは電話装置である。近年、通信方式のデジタル化に伴

い、電話装置においても各種サービスを受けることができ、然も、画像等が送れる高機能な電話装置が実用化されている。従来の電話装置において画像を伝送する場合には、例えば、現行のアナログ回線が使用され、静止画のみが伝送される。また、通信サービスにおいては、通常回線交換接続以外にも、3者同時通話が可能な接続サービスが提供されつつある。このようなサービスは、今後、さらに通信網の整備や通信方式の高度化に伴って、複雑多岐にわたるものとなることが予想され、高度な情報通信サービスの一つとして多地点間におけるテレビ会議や3人以上の話者が同時通話可能な電話会議等が考えられる。

【0003】

【発明が解決しようとする課題】しかしながら、一つの電話装置の受話器を用いて複数の人と会話しようとしても、本来の会話のように相手の方を向いて声の大きさを調節したり、顔の表情を確認しながら会話を進めることができないため、通常の会話とは感覚が異なり、円滑にコミュニケーションを図ることができない問題点があった。また、同時に会話する人数が多くなるにつれて、会話の中心となる相手の話を聞き分けることが困難になり使い勝手が悪い問題点があった。

【0004】また、多地点間におけるテレビ会議や3人以上の話者が同時通話可能な電話会議等を行う場合には、その場所毎にマイクの状態やマイクと話者との距離が異なるためにそれぞれ音量レベルが異なり、音量レベル等を全体として調整する必要が生じ、制御情報や制御系が複雑になる問題点を有していた。

【0005】従って、この発明の目的は、会話の中心となる相手の話を聞き分けることが容易で使い勝手が良く、然も、各話者の音量レベルが自動的に適正レベルとすることができる音声通話装置および音声通話システムを提供することにある。

【0006】

【課題を解決するための手段】以上の問題を解決するために、請求項1の発明は、通信を介して複数の人が音声通話するようになされた音声通話システムに用いられる音声通話装置において、通話中に各人の音声の音量を検出する音量検出手段と、上記音量検出手段により検出される音量を所定のしきい値で2値化することにより音声の有無を判別する判別手段と、上記判別手段の判別結果により、通話中の各人の音声の重なりを判定し、上記音声の重なりに応じて通話中の各人の音声のレベルを調整する音量調整手段とを備えたことを特徴とする音声通話装置である。

【0007】請求項2の発明は、通信を介して複数の人が音声通話するようになされた音声通話システムにおいて、双方向通信路を介して複数の音声通話装置が結合され、上記音声通話装置は、他の音声通話装置との通信のための通信手段と、通話中に各人の音声の音量を検出す

る音量検出手段と、上記音量検出手段により検出される音量を所定のしきい値で2値化することにより音声の有無を判別する判別手段と、上記判別手段の判別結果により、通話中の各人の音声の重なりを判定し、上記音声の重なりに応じて通話中の各人の音声のレベルを調整する音量調整手段とを備えたことを特徴とする音声通話システムである。

【0008】この発明では、制御部において、処理対象区間における各話者の音量レベルの検出がなされ、所定のしきい値で音量レベルを2値化することで発声時間が検出されると共に、重なり時間が検出される。そして、音声の重なり時間を $T1$ とし、話相手と想定される人の発声時間を $T2$ としてレベル抑制の重み W が($W = T1/T2$)により算出される。重み W が各話者の対応して設けられたゲインコントロールアンプのそれぞれに供給され、ゲインコントロールアンプが重み W により制御されることにより、会話の相手以外の話者の発声音量が抑えられると共に、自分が主に話をしている場合には、自然に他の人全体の音量が抑えられて通話環境が快適なものとなる。

【0009】

【発明の実施の形態】以下、この発明の一実施形態について図面を参照して説明する。図1は、この発明による音声通話装置によって構成される電話会議システムの一例を示し、図2は、この発明の一実施形態の全体構成を示す。なお、図1および図2において対応する箇所に関しては、同一の参照符号が付されている。

【0010】図1に示すように多地点に位置する話者A、B、C、Dの4人によって電話会議がなされるものとする。この場合においては、話者A、B、C、Dのそれぞれに対して同一の構成の音声通話装置1a、1b、1c、1dが設置される。音声通話装置1a、1b、1c、1dのそれぞれは、制御ユニット10、アンプ12、スピーカ13およびマイクロホン14により構成される。音声通話装置1a、1b、1c、1dの制御ユニット10は、出力端子11と入力端子15とを有し、出力端子11にアンプ12を介してスピーカ13が接続され、入力端子15にマイクロホンが接続される。音声通話装置1a、1b、1c、1dのそれぞれが双方向通信路としてのネットワーク/電話回線2に接続され、話者A、B、C、Dの4者間においてネットワーク/電話回線2を介して互いに通話が可能な状態とされる。

【0011】音声通話装置1a、1b、1c、1dの制御ユニット19のそれぞれは、図2に示すようにネットワークインターフェース回路21、伝送符号/復号器22、制御部23、4個のD/A変換器24a、24b、24c、24d、4個のゲインコントロールアンプ26a、26b、26c、26d、A/D変換器25および加算器27により構成されている。なお、D/A変換器およびゲインコントロールアンプは、会議システムに含

まれる人数に応じて設けられている。

【0012】制御ユニット10内の制御部23は、音声処理用のメモリ31、音量検出回路32、発生区間判定回路33、重なり区間判定回路34、重み発生部35、D/A変換器36および図示せずマイクロコンピュータ等により構成されており、会話が行われている組を自動的に検出することによって話者の音量を調節するための重み W を形成し、この重みをゲインコントロールアンプ26a~26dの制御端子に供給する。この例では、重み W は、音量を抑制するものであり、ゲインコントロールアンプ26a~26dは、重み W が大きな値ほどそのゲインが小に制御される。

【0013】話者A、B、C、Dの4者のいずれの地点の音声通話装置1a、1b、1c、1dにおいても同様であるため、話者Aの地点の音声通話装置1aを例に挙げてその構成と動作について以下に説明する。ネットワーク/電話回線2およびネットワークインターフェース回路21を介して他の話者B、C、Dの音声情報が伝送符号/復号器22に供給される。

【0014】伝送符号/復号器22において所定の伝送形態とされている音声情報が復号され、復号後の音声情報がさらに識別情報に基づいて話者B、C、Dそれぞれの音声情報D2、D3、D4に分離される。音声情報D2が制御部23の音声処理用のメモリ31に供給されると共に、D/A変換器24bに供給され、音声情報D3が制御部23の音声処理用のメモリ31に供給されると共に、D/A変換器24cに供給され、音声情報D4が制御部23の音声処理用のメモリ31に供給されると共に、D/A変換器24dに供給される。

【0015】D/A変換器24bにおいて、音声情報D2がデジタル・アナログ変換され、アナログの音声信号S2が形成される。この音声信号S2がゲインコントロールアンプ26bに供給される。同様にD/A変換器24cおよび24dにおいて、音声情報D3およびD4のそれぞれがデジタル・アナログ変換され、アナログの音声信号S3およびS4が形成される。この音声信号S3がゲインコントロールアンプ26cに供給され、音声信号S4がゲインコントロールアンプ26dに供給される。

【0016】一方、マイクロホン14が接続されている入力端子15には、自分(話者A)の音声信号が供給される。入力端子15を介して話者Aの入力音声信号がA/D変換器25に供給される。A/D変換器25において、入力音声信号がアナログ・デジタル変換され、音声情報D1が形成される。この音声情報D1が伝送符号/復号器22に供給される。伝送符号/復号器22において、音声情報がさらに所定の伝送形態に符号化され、この出力がネットワークインターフェース回路21およびネットワーク/電話回線2を介して接続地点のそれぞれに対して送出される。また、音声情報D1は、伝送符

号化されることなく、伝送符号／復号器22を介して制御部23の音声処理用のメモリ31に供給されると共に、D/A変換器24aに供給される。D/A変換器24aにおいて、音声情報D1がデジタル・アナログ変換され、アナログの音声信号S1が形成される。この音声信号S1がゲインコントロールアンプ26aに供給される。

【0017】ゲインコントロールアンプ26a、26b、26c、26dのそれぞれは、制御端子を有しており、この制御端子に供給される制御信号（重みW）に応じてそれぞれの信号レベルを調整して出力する。ゲインコントロールアンプ26aの制御端子には、制御部23からの制御信号C1が供給されており、音声信号S1が所定レベルに制御されて加算器27の第1の入力端子に供給される。同様にゲインコントロールアンプ26bの制御端子には、制御部23からの制御信号C2が供給され、ゲインコントロールアンプ26cの制御端子には、制御部23からの制御信号C3が供給され、ゲインコントロールアンプ26dの制御端子には、制御部23からの制御信号C4が供給される。従って、音声信号S2が所定レベルに制御されて加算器27の第2の入力端子に供給され、音声信号S3が所定レベルに制御されて加算器27の第3の入力端子に供給され、音声信号S4が所定レベルに制御されて加算器27の第4の入力端子に供給される。

【0018】加算器27において、ゲインコントロールアンプ26a、26b、26c、26dのそれぞれの出力が加算され、合成音声信号が形成される。この合成音声信号が出力端子11を介して取り出され、スピーカ13が接続されたアンプ12に供給される。後述するように、制御部23において、各人の音声信号のレベルを制御することによって、スピーカ13により再生される音声は、会話の組以外の話者（すなわち、会話のタイミングが合っていない話者）の発声音量が抑えられると共に、自分が主に話をしている場合には、自然に自分以外の話者の音量が抑えられたものとされる。なお、他の地点の音声通話装置1b、1c、1dにおいても同様に処理がなされ、各地点においてそれぞれに快適な会話環境が形成される。

【0019】上述した制御部23の動作について詳細に説明する。通常、多人数で会話をする時には、自分の発声と重なるように話す人の声は煩わしく、また、それでも自分の話を聞かせたい場合には、重なって発声している相手よりも大きな声で発声するように心掛けるものであり、逆に、聞きたい相手の発声中には、自分の発声を控えるものである。さらに、聞こえる音量が少々低くても、興味のある話題には、敏感に反応するという面もある。このように会話時の発声音量は、円滑にコミュニケーションを図る上で重要な要素である。この発明は、制御部23の処理によって、各話者の音声レベルを制御

し、自然なユーザーインターフェースを実現しようとするものである。

【0020】制御部23の音声処理用のメモリ31には、上述したように伝送符号／復号器22からの音声情報D1、D2、D3、D4が供給されており、逐次音声情報D1、D2、D3、D4がメモリ31に蓄積される。メモリ31の蓄積時間、即ち、処理対象区間Ctは、必要に応じて可変することができ、例えば、(Ct=30sec)に設定されている。つまり、メモリ31において、現時刻から30sec前までの区間の音声情報D1、D2、D3、D4のそれぞれがメモリ31に書き込まれて一時的に保持される。この処理対象区間Ctで処理を行うことで、制御部23の反応が必要以上に過敏になることが防止される。なお、この説明においては、処理対象区間Ctが予め設定されたものであるが、処理対象区間Ctを可変するようにしても良い。メモリ31に蓄積された音声情報D1、D2、D3、D4のそれぞれが読み出され、音量検出回路32に供給される。

【0021】音量検出回路32において、各話者A、B、C、Dの音量レベルの検出がなされ、音量検出回路32の出力が発声区間判定回路33に供給される。発声区間判定回路33は、所定のしきい値で音量レベルを2値化することで発声区間を判定すると共に、処理対象区間Ct内の発声時間をカウントする。発声区間判定回路33において、音声レベルが所定値以上である場合には、例えば、ハイレベルとされる発声区間情報と発声時間を示す情報とが形成される。各話者A、B、C、Dのそれぞれの発声区間情報と発声時間を示す情報とが重なり区間判定回路34に供給される。

【0022】重なり区間判定回路34は、例えば、発声区間判定回路33からの各話者A、B、C、Dのそれぞれの発声区間情報の論理積により重なり区間を判定し、その重なり時間をカウントする。重なり区間判定回路33において、発声区間の重なり合う関係を示す情報と、重なり時間を示す情報とが形成され、この情報が重み発生部35に供給される。また、音声区間判定回路33からの発声時間を示す情報がそのまま重み発生部35に供給される。

【0023】重み発生部35において、所定の演算による重みWが生成される。例えば、音声の重なり時間をT1とし、話相手と想定される人の発声時間をT2とした場合には、下式に示される演算により重みWが算出される。なお、予め重みWの最大値Wmaxが設定されており、重みWが(W<Wmax)とされる。

【0024】 $W = T1 / T2$ (W<Wmax)

話者A、B、C、Dの各組み合わせでそれぞれに重みWが算出され、得られた重みWに応じた制御情報が形成される。この制御情報がD/A変換器36に供給される。D/A変換器36において、重み発生部35からの制御情報がデジタル・アナログ変換され、各話者A、B、

C, Dの音声レベルを制御する4つの制御信号C1, C2, C3, C4が形成される。

【0025】この制御信号C1, C2, C3, C4のそれぞれが前述したゲインコントロールアンプ26a, 26b, 26c, 26dのそれぞれに供給される。ゲインコントロールアンプ26a, 26b, 26c, 26dのそれぞれが制御信号C1, C2, C3, C4により制御されることにより、会話のタイミングが合っていない話者の発声音量が抑えられると共に、自分が主に話をしてしている場合には、自然に他の人全体の音量が抑えられる。また、逆に発声タイミングを合わせることで話相手以外の他の話者の音量を抑えることができる。

【0026】なお、上述した制御部23の動作は、自動的になされるものであるが、例えば、制御部23が手動モードに設定された場合には、各話者A, B, C, Dに対する重みWをそれぞれに設定して手動で調整することが可能とされている。

【0027】上述した制御部23の動作に関してさらに具体的な会話例を挙げて詳細に説明する。各話者A, B, C, Dの全員が会話に参加しており、例えば、図3で示すように発声区間が検出されているものとする。なお、図3においては、横軸が時間の経過を示し、最上段から話者A, B, C, Dの発声区間が示されている。また、図3において矢印51で示される数値が発声区間判定回路33において、カウントされた処理対象区間Ct内の各話者A, B, C, Dの発声時間である。

【0028】重なり区間判定回路34において、発声区間判定回路33からの各話者A, B, C, Dのそれぞれの発声区間情報の論理積により重なり区間が判定され、その重なり時間がカウントされる。図3に示すタイミングで会話がなされている場合には、図4に示すように各話者A, B, C, Dの全ての組み合わせで重なり時間T1が検出される。なお、図4において、各数値の単位は秒であり、()で括弧で表示されている数値は、発声区間判定回路33において、カウントされた処理対象区間Ct内の各話者A, B, C, Dの発声時間である。

【0029】各話者A, B, C, Dの全ての組み合わせで重なり時間T1が検出されると、次に、重み発生部35において、各話者A, B, C, Dの全ての組み合わせで重みWが算出される。重みWを図5に示す。例えば、話者Aの地点において話者Bに対する重みWを算出する場合には、重なり時間T1が2.5秒であり、話者Bの発声時間T2が10秒であるため、 $(2.5/10 = 0.25)$ が算出される。また、話者AとDとの重みWを算出する場合には、重なり時間T1が12.5秒であり、話者Dの発声時間T2が25秒であるため、 $(12.5/25 = 0.5)$ が算出される。同様に各話者A, B, C, Dの全ての組み合わせで重みWが算出される。なお、重み付けを行う場合においては、予め重みWの最大値Wmaxが0.9に設定されており、それによっ

て、会話に支障をきたさない範囲での音量レベルの制御が可能とされる。

【0030】そして、図5に示す重みWに応じてアナログの制御信号C1~C4がD/A変換器36から発生する。制御信号C1~C4によって、重みWが大きいほどゲインが小さくなるように、ゲインコントロールアンプ26a, 26b, 26c, 26dのそれぞれのゲインが制御される。

【0031】なお、上述した一実施形態の説明においては、発声区間情報の論理積により重なり区間の判定を行って重なり時間T1を検出する場合について説明したが、例えば、重なっていない区間を検出しても演算により重なり時間T1を算出することができ、重なり時間T1の検出方法としては、種々のものが可能である。

【0032】さらに、上述した一実施形態の説明においては、話者が4人の場合について説明したが、当然のことながら、3人以上で4人以外の場合においてもこの発明を適用することができ、この場合においては、D/A変換器およびゲインコントロールアンプのそれぞれが対応して増設される。

【0033】

【発明の効果】この発明では、制御部が設けられ、制御部において、処理対象区間Ctにおける各話者の音量レベルの検出がなされ、所定のしきい値で音量レベルを2値化することで発声時間が検出されると共に、重なり時間が検出される。そして、音声の重なり時間をT1とし、話相手と想定される人の発声時間をT2として重みWが $(W = T1/T2)$ により算出される。従って、この発明に依れば、会話のタイミングが合っていない話者の発声音量が抑えられると共に、自分が主に話をしてしている場合には、自然に他の人全体の音量が抑えられて円滑にコミュニケーションを図ることができる。

【図面の簡単な説明】

【図1】この発明による音声通話装置を用いた電話会議システムのブロック図である。

【図2】この発明の一実施形態の主要な部分の構成を示すブロック図である。

【図3】この発明の一実施形態の動作説明に用いる略線図である。

【図4】この発明の一実施形態の動作説明に用いる略線図である。

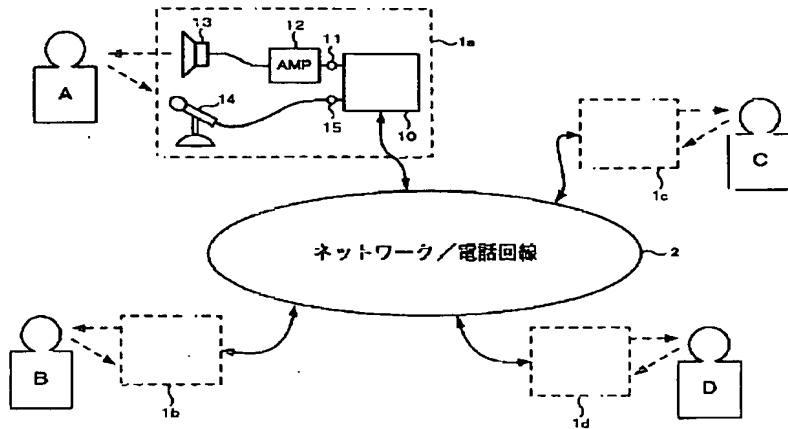
【図5】この発明の一実施形態の動作説明に用いる略線図である。

【符号の説明】

1a, 1b, 1c, 1d・・・音声通話装置、2・・・ネットワーク/電話回線、10・・・制御ユニット、21・・・ネットワークインターフェース回路、23・・・制御部、26a, 26b, 26c, 26d・・・ゲインコントロールアンプ、31・・・音声処理用のメモリ、32・・・音量検出回路、33・・・発声区間判定

回路、34・・・重なり区間判定回路、35・・・重み発生部 発生部

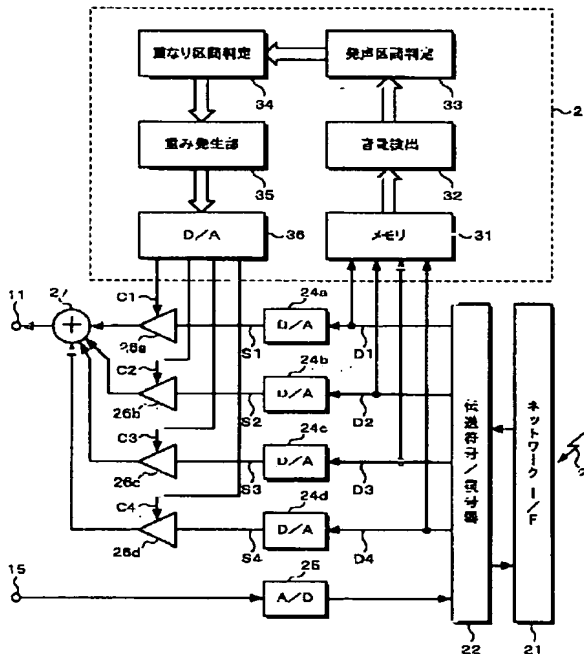
【図1】



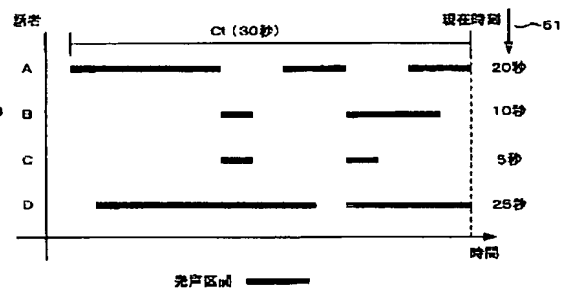
【図4】

話者（発声の秒数）／相手	A	B	C	D
A (20)	—	2.5	0	12.5
B (10)	2.5	—	5	10
C (5)	0	5	—	5
D (25)	12.5	10	5	—

【図2】



【図3】



【図5】

話者／相手	A	B	C	D
A	—	0.25	0	0.5
B	0.125	—	0.9	0.4
C	0	0.5	—	0.2
D	0.625	0.9	0.9	—

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.